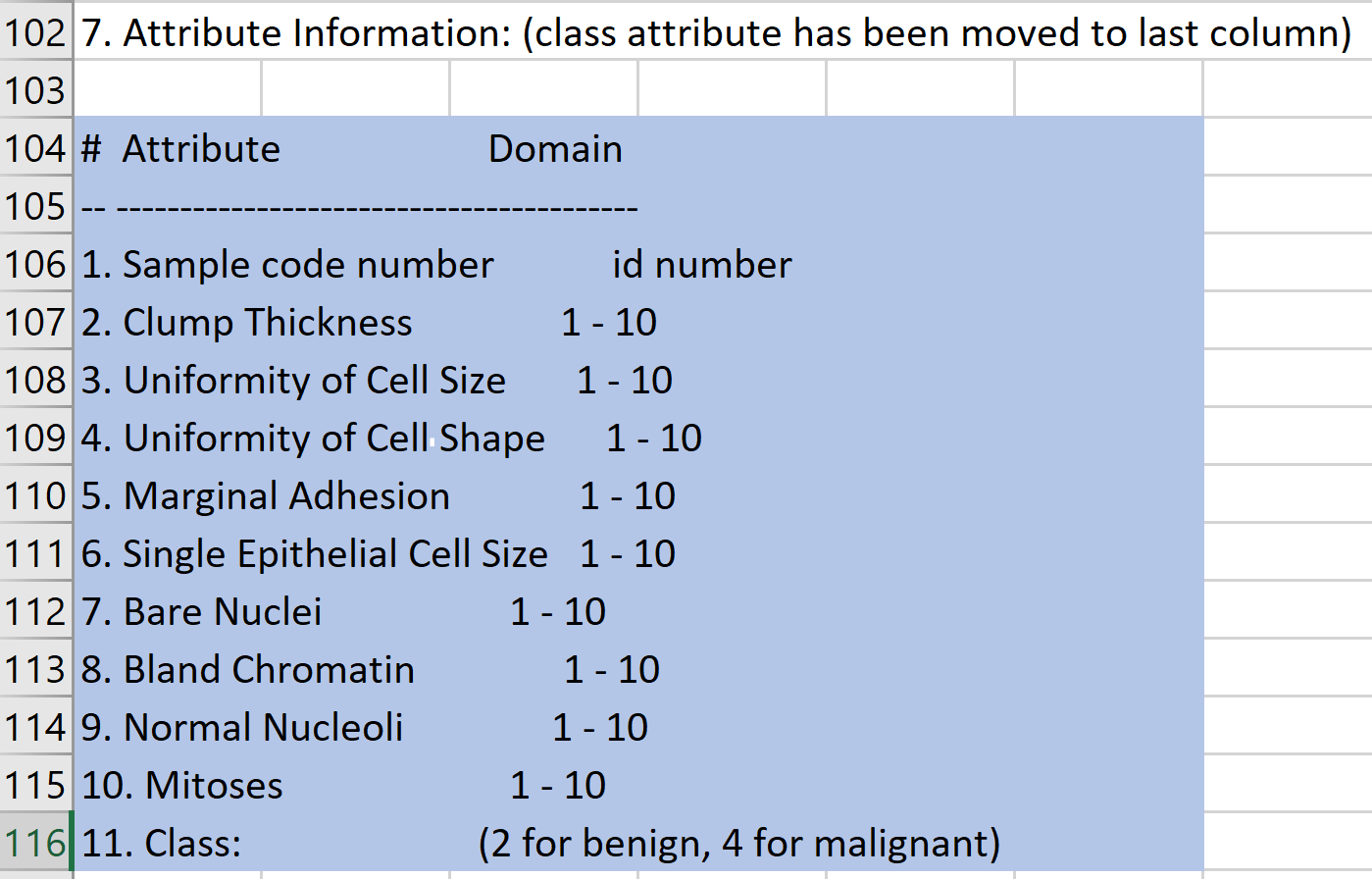
Classification Exercises

Datasets are either in data folder or in Matlab’s demo dataset directory.

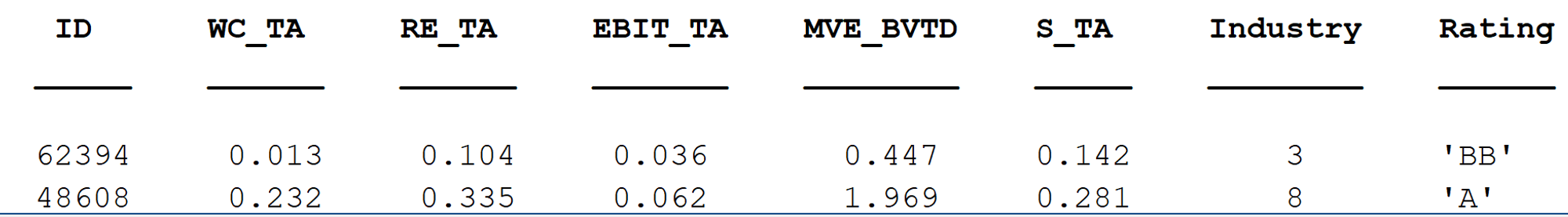
Use Matlab’s 5-fold cross validation (its default) as the model evaluation method.

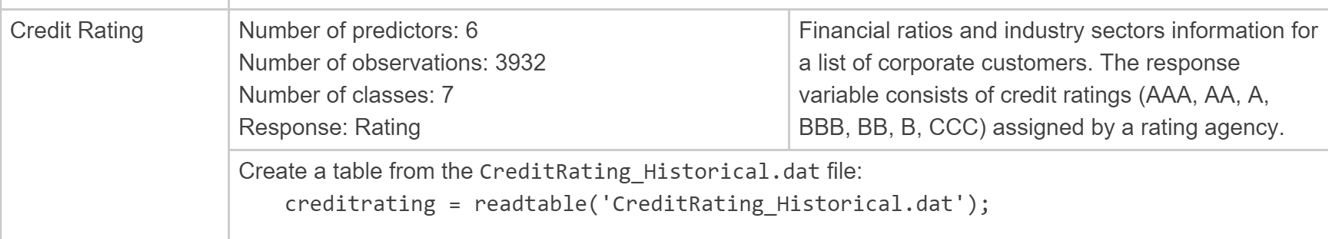
1. Breast cancer data
   1. Use the dataset “Breast Cancer Wisconsin.xls” to build a model to diagnose breast cancer.
   2. The Description sheet provides a description of the data. Row 104+ contains description of the columns.



* 1. Build a classification model using various methods below and compare your models.
     1. KNN .. Try a few K values. Report your best K
     2. Logistic Regression
     3. Decision Tree.. Try a few “Max Number of Splits”. Report your best choice.
     4. SVM.
     5. Ensemble: Boosted Tree and Bagged Tree.

1. Handwritten digits recognition
   1. MNIST2 (Matlab mat file) in the Datasets directory is a version of the MNIST handwritten digits data. Each image is represented by 20x20 greyscale array that has been transformed to a 400 element vector. The last column represents the labeled classification of 0-9 for each image.
   2. Use the data to train a classification model to recognize the handwritten digits. Choose your best model and report accuracy and confusion matrix.
   3. Later you can use the same data and build a Neural Net model to compare.
2. Credit Rating Data
   1. Load the Credit Rating Historical data from Matlab’s demo datasets using the instruction below.
   2. Note that 1st column is ID, the 7th column is industry group. The last column is one of the 7 ratings (AAA🡪CCC)
   3. Build a classification model without the ID and industry group to generate credit ratings of companies. Report accuracy and Confusion Matrix. (If you would like to include industry group you need to one-hot convert this column).
   4. Later you can use the same data and build a Neural Net model to compare.





Problem #4 is Optional

1. Rerun IRIS data
   1. Load the IRIS data from Matlab’s demo datasets.
   2. Build the model using various methods. You can also look into the step-by-step Decision Tree example in the Matlab link (not here, need to go to website first)

